R E P O R T R E S U M E S

ED 019 978 LI 000 583

FINAL REPORT OF THE RESEARCH PROJECT FOR THE EVALUATION OF THE UDC AS THE INDEXING LANGUAGE FOR A MECHANIZED REFERENCE RETRIEVAL SYSTEM.

BY- FREEMAN, ROBERT R. ATHERTON, PAULINE
AMERICAN INST. OF PHYSICS, NEW YORK, N.Y.

REPORT NUMBER AIP-UDC-9
EDRS PRICE MF-$0.25 HC-$1.32 31P.

DESCRIPTORS- *CLASSIFICATION, *INFORMATION RETRIEVAL, *INDEXING, *INFORMATION STORAGE, INFORMATION SYSTEMS, AUTOMATION, FEASIBILITY STUDIES, DOCUMENTATION, TESTING, UNIVERSAL DECIMAL CLASSIFICATION (UDC)

THE BACKGROUND, OBJECTIVES, AND ACCOMPLISHMENTS OF THE PROJECT ARE REVIEWED BRIEFLY. SPECIFIC AREAS DISCUSSED ARE ENGLISH-LANGUAGE UDC SCHEDULES, A COMPUTER-BASED UDC FILE MANAGEMENT SYSTEM, DATA BASES FOR RETRIEVAL EXPERIMENTS, AND RETRIEVAL SYSTEM EVALUATION. THE CONCLUSIONS DEAL WITH THE USEFULNESS OF THE UDC FOR MECHANIZED RETRIEVAL SYSTEMS, NEEDED RESEARCH, NEEDED ORGANIZATIONAL EFFORT, AND A PROPOSED INTERNATIONAL SEMINAR. APPENDIX 1 GIVES A BIBLIOGRAPHY OF AIP/UDC PROJECT REPORTS WITH ABSTRACTS. APPENDIX 2 GIVES INFORMATION ABOUT THE AIP/UDC PROJECT MAGNETIC TAPES AND OTHER TAPES CONTAINING AIP/UDC PROJECT MATERIALS. THESE MATERIALS INCLUDE ABOUT 50,000 UDC-INDEXED DOCUMENT REFERENCES AND THE UDC ENGLISH LANGUAGE MASTER FILES. APPENDIX 3 GIVES A SUMMARY OF THE STATUS OF THE UDC IN ENGLISH. APPENDIX 4, ISSUED SEPARATELY AS LI 000 584, IS THE UDC ENGLISH MEDIUM EDITION MANUSCRIPT. APPENDIX 5, ALSO ISSUED SEPARATELY AS LI 000 585, CONTAINS THE UDC ENGLISH LANGUAGE MASTER FILES. (AUTHOR/RP)

ED019978

LI000583

FINAL REPORT
of the
RESEARCH PROJECT FOR THE EVALUATION OF THE UDC AS THE INDEXING LANGUAGE FOR A MECHANIZED REFERENCE RETRIEVAL SYSTEM

by
Robert R. Freeman and Pauline Atherton

AMERICAN INSTITUTE OF PHYSICS
UDC PROJECT
REPORT NO. AIP/UDC-9
National Science Foundation Grant GN-433

LI000583

AIP/UDC Project Reports

1. Freeman, Robert R., Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System: An Introduction, Report AIP/DRP UDC-1, October 1, 1965.

2. Freeman, Robert R., Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System: Progress Report for the Period July 1, 1965 - January 31, 1966, Report AIP/DRP UDC-2, February 1, 1966.

3. Freeman, Robert R., Modern Approaches to the Management of a Classification, Report AIP/UDC-3, October 1, 1966.

4. Russell, Martin and Freeman, Robert R., Computer-Aided Indexing of a Scientific Abstracts Journal by the UDC with UNIDEK: a Case Study, Report AIP/UDC-4, April 1, 1967.

5. Freeman, Robert R. and Pauline Atherton, File Organization and Search Strategy Using the Universal Decimal Classification in Mechanized Reference Retrieval Systems, Report AIP/UDC-5, September 15, 1967.

6. Freeman, Robert R., Evaluation of the Retrieval of Metallurgical Document References Using the Universal Decimal Classification in a Computer-Based System, Report AIP/UDC-6, April 1, 1968.

7. Freeman, Robert R. and Pauline Atherton, AUDACIOUS - an Experiment with an On-Line, Interactive Reference Retrieval System Using the Universal Decimal Classification as the Index Language in the Field of Nuclear Science, Report AIP/UDC-7, April 25, 1968.

8. Atherton, Pauline, Donald W. King, and Robert R. Freeman, Evaluation of the Retrieval of Nuclear Science Document References Using the Universal Decimal Classification in a Computer-Based System, Report AIP/UDC-8 May 1, 1968.

9. Freeman, Robert R. and Pauline Atherton, Final Report of the Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System, Report AIP/UDC-9, May 1, 1968.

Copies of the reports listed above may be obtained from the Information Division, American Institute of Physics, 335 East 45th Street, New York, New York 10017. The reports are also available from the Clearinghouse for Federal Scientific and Technical Information, 5285 Port Royal Road, Springfield, Virginia 22151 at $0.65 for microfiche or $3.00 for full-size copy.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

of the

RESEARCH PROJECT FOR THE EVALUATION OF THE UDC AS THE INDEXING LANGUAGE FOR A MECHANIZED REFERENCE RETRIEVAL SYSTEM

by
Robert R. Freeman and Pauline Atherton

AMERICAN INSTITUTE OF PHYSICS
UDC PROJECT
REPORT NO. AIP/UDC-9
National Science Foundation Grant GN-433

Abstract

Freeman, Robert R., and Pauline Atherton, Final Report of the Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System, Report AIP/UDC-9 under NSF Grant GN-433. New York, American Institute of Physics, May 1, 1968.

The background, objectives, and accomplishments of the project are reviewed briefly. Specific areas discussed are English-language UDC schedules, a computer-based UDC file management system, data bases for retrieval experiments, batch-process and on-line, interactive information retrieval systems, and retrieval system evaluation. The conclusions deal with the usefulness of the UDC for mechanized retrieval systems, needed research, needed organizational effort, and a proposed international seminar. Several appendices summarize the current state of the UDC in English and the availability of magnetic tape and microfilm files developed by the project.

UDC 025.3+025.45UDC+651.83.012.1

Explanation of UDC numbers:

025.3	Cataloging and indexing - Information retrieval systems
025.45UDC	Decimal classifications - UDC
651.83	Indexing and retrieval methods
.012.1	Experimental testing and evaluation

FINAL REPORT OF THE RESEARCH PROJECT FOR THE EVALUATION OF THE UDC AS THE INDEXING LANGUAGE FOR A MECHANIZED REFERENCE RETRIEVAL SYSTEM

by

Robert R. Freeman and Pauline Atherton

Contents

Section	Title	Page
	Abstract	ii
1.	Introduction	1
2.	Objectives	2
3.	Summary of results	2
3.1.	UDC Schedules	4
3.2.	The UDC File Management System	4
3.3.	Data bases for experimentation	4
3.4.	Batch Process retrieval system	6
3.5.	Interactive retrieval system	6
3.6.	Retrieval system evaluation	6
4.	Conclusions and recommendations	7
4.1.	Usefulness of UDC	7
4.2.	Needed research	10
4.3.	Needed organizational effort	12
4.4.	Seminar on UDC in a mechanized retrieval system	13
5.	Acknowledgements	14

Appendices

1. Bibliography of AIP/UDC Project Reports, with Abstracts
2. AIP/UDC Project Magnetic Tapes and Other Tapes Containing AIP/UDC Project Materials
3. Summary of the Status of the UDC in English
4. UDC English Medium Edition Manuscript
5. The UDC English Language Master Files

Final Report of the Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System

by

Robert R. Freeman and Pauline Atherton

1. Introduction. The development of modern data processing equipment and techniques, beginning about two decades ago, coincided with a crisis in the handling of scientific information by libraries and information centers. Increasingly large and rapidly growing document collections and changing user needs, brought on by the burgeoning of scientific research, placed difficult stresses on existing science information systems.

The Universal Decimal Classification (UDC) had been developed and promulgated since the beginning of the twentieth century, as a means of classifying and indexing documents in any field of knowledge, but especially in science and technology. While UDC gained wide acceptance, especially in Europe, the questions of whether it could be used in newer, mechanized information systems and whether it could be kept up-to-date in the face of rapid change were naturally raised.

In the United States, The National Science Foundation supported pioneering attempts to answer these questions, begun as early as 1961 by Malcolm Rigby at the American Meteorological Society (AMS). These efforts led to the demonstration of techniques for computer handling of UDC schedules and the preparation of computer-printed systematic indexes based on UDC. An experimental current-awareness service, Meteorological and Geoastrophysical Titles, incorporated this type of index, for which the name UNIDEK was coined. AMS has continued to produce UNIDEK indexes of the accessions of the U. S. National Oceanographic Data Center, while the American Geological Institute also adopted the techniques for annual indexes to Geoscience Abstracts.

By 1965, it still remained to explore and demonstrate the use of UDC in a mechanized retrieval system and to cope with the problem of managing

the UDC itself, which had never reached the stage of full publication in the English language. The UDC Project of the American Institute of Physics (AIP) was created by the National Science Foundation (NSF) to address these problems.

2. Objectives. The objectives of the project were given in the proposal submitted by AIP to NSF, as follows:

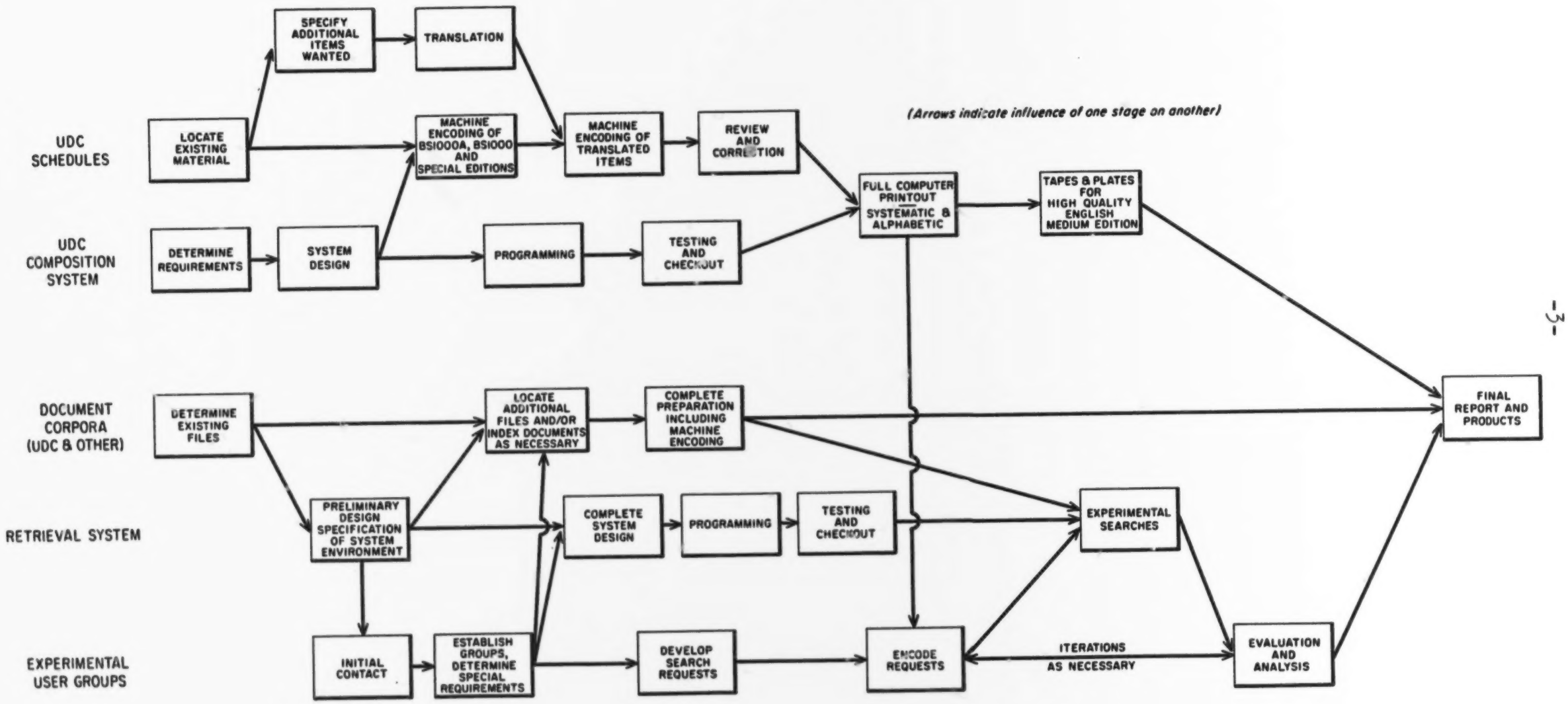
> "We propose to design and demonstrate a reference retrieval system in which the coding subsystem is the UDC and the display and search subsystem is at least partially mechanized. The objective is to evaluate the ability of the UDC to present relevant references to the user in a useful display and to screen out the irrelevant references. The design of the retrieval tests using this mechanized UDC system and other mechanized systems will be carefully constructed in order to insure comparable results and proper assessment of relevance by user group representatives. A proposed standard description for evaluation tests will be followed.
>
> The experimental system proposed here is intended to demonstrate the capabilities of the UDC as an indexing language as it is being applied in real life situations. As a part of the feedback from this study, we would expect to learn about areas in which the UDC might be improved with regard either to conceptual relations or to notational devices when it is used in a mechanized system."

The plan of work for accomplishing these objectives is shown in Figure 1. The following section describes the results of these steps.

3. Summary of Results. The results are described in terms of major areas of accomplishment. They are (1) the UDC schedules, (2) the UDC File Management system, (3) the data bases for retrieval system experimentation, (4) the batch process retrieval system, (5) the interactive retrieval system, and (6) retrieval system evaluation. Each of these areas is described in detail in reports published by the project or in appendices of this report. Consequently, only brief summaries with references are presented here.

FIGURE 1

RESEARCH PROJECT FOR THE EVALUATION OF THE UDC AS THE INDEXING LANGUAGE IN A MECHANIZED REFERENCE RETRIEVAL SYSTEM



R.R.F. 8/24/65

3.1. UDC Schedules. English language schedules from both published editions and manuscript sources were entered into a single machine-readable file. The entire UDC is represented to at least the level of the English Abridged Edition, while approximately 93% is represented to at least the level of a Medium Edition. Appendix 3 shows the UDC classes which are present in the file. Appendix 5 gives details of the availability of the complete master file on both magnetic tape and microfilm.

By special arrangement with the International Federation for Documentation (FID), a significant beginning was made toward the development of an English Medium Edition of UDC. The state of this effort and the arrangements for availability of the manuscript are described in Appendix 4.

3.2. The UDC File Management System. Building on techniques developed by AMS, we developed a computer-based file management system, which was used to maintain the master file referred to above. The program also provides for automatic generation of a keyword index to the schedules and for input to a photocomposition system. With the help of a subcontractor, we were able to demonstrate that high-quality composition of classification schedules directly from a machine-readable file is possible. The file management system and techniques, including photocomposition are discussed in project reports 2 and 3.

3.3. Data bases for experimentation. Six files of UDC-indexed document references were created and converted to magnetic tape for use in the experimental retrieval systems described in sections 3.4. and 3.5. Tape numbers, contents, and availability are described in Appendix 2. The files consist of the following data.

3.3.1. Nuclear science. The U.S. Atomic Energy Commission provided a tape containing the descriptive cataloging and Euratom Keyword indexing data for Nuclear Science Abstracts, vol. 19, no. 11, June 15, 1965. The file contained 2,330 references.

A team of experienced UDC indexers was assembled from among the staff of several installations of the United Kingdom Atomic Energy Authority (UKAEA), under the supervision of Mr. Jack Terry. This team provided UDC-indexing records for each document reference in the collection. These records were then merged into the file. Indexing was done according to the Special Subject Edition of UDC for Nuclear Science and Technology, augmented by a code of practice developed by UKAEA.

3.3.2. Geology. A file of 20,892 document references, consisting of abstract numbers, titles, and UDC numbers, from Geoscience Abstracts, volumes 6-8 (1964-1966), was contributed in machine-readable form by the American Geological Institute.

3.3.3. Meteorology. A file of approximately 9,000 document references, consisting of authors, subject headings, and abstract numbers, was contributed in machine-readable form by the American Meteorological Society. The data were from the 1965 volume of Meteorological and Geoastrophysical Abstracts (MGA). The AIP/UDC Project added UDC numbers from the printed MGA to complete the file.

3.3.4. Oceanography. A file of 3,900 document references, consisting of authors, titles, references, and UDC numbers was contributed in machine-readable form by the American Meteorological Society. The data represent the 1966 and January-March, 1967 accessions of the National Oceanographic Data Center.

3.3.5. Antarctic Studies. A file of 4,000 document references, consisting of abstract numbers, titles, and UDC numbers was created from the Antarctic Bibliography, published by the Library of Congress under the sponsorship of the Office of Antarctic Programs of the National Science Foundation. The data represent abstracts originally published from 1962-1966.

3.3.6. Metallurgy File. A file of 9,159 document references, consisting of abstract numbers, titles, and UDC numbers, was created from a card file of abstracts contributed by the Iron and Steel Institute, located in London. The data were originally published as the 1965 coverage of the Abstract and Book Title Card Service (ABTICS) of that organization.

3.4. Batch Process Retrieval System. An existing package of computer programs for the IBM 1401, known as the Combined File Search System, served as the mechanism for testing and evaluation of the UDC, as described in section 3.6. The details of the operation of this retrieval system were presented in project report number 5.

3.5. Interactive Retrieval System. The feasibility of use of UDC in an on-line, interactive retrieval system was demonstrated with a system developed by a subcontractor, using the nuclear science data base (see 3.3.1.) and the Special Subject Edition of UDC for Nuclear Science and Technology. The results of the use of this system, referred to as "AUDACIOUS" (Automatic Direct Access to Information with the On-Line UDC System), are documented in project report number 7.

3.6. Retrieval System Evaluation. Several experiments were undertaken for the purposes of demonstrating the use of UDC in a computer-based retrieval system and evaluating the degree of success which might be expected for

such applications. For the field of metallurgy, tests performed with the cooperation of the Iron and Steel Institute (London) and the American Society for Metals are reported in project report number 6. In the field of nuclear science, tests performed in cooperation with the U. S. Atomic Energy Commission, Euratom, and UKAEA are reported in project report number 8.

4. Conclusions and Recommendations.

4.1. Usefulness of UDC. The purpose of the work of this project has been to bring the light of practical experience to bear on an area in which there had previously been only speculation, thus providing guidance for the information system planner who is faced with deciding what role, if any, the UDC should play in a particular system.

There is no longer any doubt that the UDC can be used as the indexing language in a mechanized system. No barriers exist to the successful use of the UDC in either a batch-processing or interactive mode.

The results of the project should lend support and encouragement to those who will consider use of UDC in computer-based retrieval systems. No insoluble problems were found, but the long-existent matter of the theory according to which the UDC will be developed in the future is seen to be accentuated by the requirements and capabilities of computer-based systems.

On the basis of experiments in a test environment which reasonably simulates a real information system, we feel justified in encouraging those who wish to make use of UDC as the indexing language in a computer-based retrieval system. To the extent that the observed results are reliable, valid, and indicative, the operating characteristics of the experimental

batch-process system are surprisingly good, especially when one recalls that the indexing for all files described in section 3.3. was done with a purely manual system in mind.

The results, particularly the failure analyses, revealed some points which should be seriously considered by system designers and managers who intend to use UDC as the indexing language in their system. These points, which may be grouped as (1) search strategies and predictive tools, (2) hierarchical searching, (3) new indexing policies, and (4) revisions and innovations in the UDC, are discussed in project report number 6.

AUDACIOUS was, to the best of our knowledge, the first on-line interactive retrieval system in which one of the widely used traditional classification and indexing tools was used. While the UDC was the tool in this case, the success of the experiment may be generalizable to other tools, such as the Dewey and Library of Congress Classifications.

For system designers, clearly, the most important implication of the results of AUDACIOUS is the need for careful consideration of the user viewpoint in all facets of the design of an interactive retrieval system. A system which is a technical success can fail to impress an information system user in many areas, some of which we have discussed in project report number 7.

In general, system planners will want to consider the UDC if there are compelling reasons. Several such reasons might be mentioned here. (1) An organization, through many years of use, may have built up large files and a skilled staff based on the use of UDC. (2) The ability to use UDC could save the not-insignificant cost of developing an indexing language. (3) The idea of an internationally used indexing language may have appeal for

organizations for whom international exchange of materials in several natural languages is important. In conditions such as these, no barrier exists to the successful use of UDC in a mechanized retrieval system.

Improvements during the past few years in communications and computer technologies strongly indicate that networks of libraries and information centers, whose resources are linked electronically, will be feasible in the not-too-distant future. Users will be able to conduct searches by means of a dialogue with the system, with access to distant as well as geographically-nearby files of information. Such networks need not be confined within national borders; Dubon[1], for example, has outlined a possible European Information Network, in which various national centers, each specializing in a given subject area, would exchange information cooperatively.

The very concept of an international network raises the question: what manner of indexing would serve adequately for users who do not share a common natural language? One solution is to use the language in which the largest volume of literature is written, i.e. English. This solution undoubtedly serves well in a situation in which the user must submit his question through an intermediary analyst who is skilled in both the subject matter and in English. However, it is open to question whether the average non-native speaker, even though he may be able to converse with another person in English, would be able to carry on a successful dialogue with a computer-based information file.

Another solution might be to make use of a form of indexing that is not dependent on natural language -- which suggests the UDC. Without

---

1. R.J. Dubon, "Implementation of an International Information Retrieval Center", pp. 339-346 in Progress in Information Science and Technology: Proceedings of the American Documentation Institute, Vol. 3, Santa Monica, California, Adrianne Press, 1966.

commenting on the present adequacy of the UDC, it should be emphasized again that tables for conversion between UDC and natural language already exist for some sixteen languages[2] and that UDC-indexing appears with original research papers in possibly hundreds of journals published in many countries and languages. Kepple[3] has pointed out the advantage of the UDC as a tool for an international library because it is not language-dependent.

A third solution, requiring greater effort to implement it, would be to permit indexing and searching to be done using a controlled natural-language vocabulary of local choice. A part of the system would then be a table of equivalences between the UDC and the natural language vocabulary. The result would be to take advantage of the hierarchical notation of the UDC without even requiring that the user be familiar with the UDC. In addition, since the UDC would be the internal form of indexing, users in any center could direct queries to the file, without regard to the original language in which the indexing was done.

4.2. Needed Research. A critical need continues to be methods for evaluation of information retrieval systems, especially those which may be employed during the design of a system in order to maximize the chances of its success according to some criteria. At the time this project commenced work, quantitative performance measures applied in a post hoc fashion reflected the state-of-the-art.

Since that time, there has been progress in three distinct areas. First, the techniques of statistical sampling and inference are coming to be applied

---

2. For a list of UDC editions, see International Federation for Documentation, FID Publications Catalogue 1968, FID 427, The Hague, January, 1968.

3. R.R. Kepple, "Serving Readers in a Special International Library", College and Research Libraries, 28(3), 203-207, 216 (1967).

to information systems so as to enable predictions of performance to be made before full-scale development of a system. Second, there has been the development of evaluation models which take into account, at least qualitatively, the behavioral factors involved in the act of judging the relevance of documents retrieved by a system. Finally, there is increasing realization that behavioral factors of man-system relationships may be more significant than quantitative performance in the evaluation of information systems.

We have attempted to make use of these advances, if only crudely at times, in compiling the results of this project's work. However, feeling keenly the lack of sufficient and integrated methods, we strongly urge that further work along this line be carried out. Methods for the design and evaluation of systems in which the user interacts directly with a machine-stored index or document file are especially needed.

We have presented a detailed discussion of the difficulties which arise at the time UDC-indexed files are searched by machine which are attributable to the structure and class definition system of the UDC. The problems of revisions and innovations reflect a deep-rooted question for the International Federation for Documentation: can the UDC be universal in the sense of being applicable to all types of information systems? Are the requirements of organizations which will use the UDC for the purpose of systematic single-entry document file organization (e.g. conventional libraries) compatible with those of organizations which will offer services based on deep indexing, highly specific questions, and the use of the computer as an aid?

From the point of view of the latter type of system, continued research into the revision of UDC according to principles and techniques of faceted

classifications seems to be indicated. We also recommend the testing of more sophisticated devices for coding syntagmatic relationships, such as the schema of relators suggested by Perreault.

Another suggestion we would make for future investigation is to explore the use of UDC in conjunction (rather than in parallel, as we did) with a suitably detailed thesaurus. UDC might be used to rapidly narrow the portion of the file to be searched to a small size, the thesaurus then being used for detailed interaction with that subset of the file. The problem of user preferences for natural language versus a numeric or other code also needs to be investigated.

Finally, our experience with AUDACIOUS (project report number 7) and several other similar systems points to a need for research on methods of teaching the use of on-line, interactive retrieval systems. The similarities of many aspects of this type of retrieval system with computer-aided instruction (CAI) suggest that the latter may provide a fruitful avenue of exploration toward a solution of this problem.

4.3. Needed Organizational Effort. Probably the most frequently-heard criticism of UDC is that it is not up-to-date in its coverage of the rapidly evolving terminology of various technical subject areas. The problem has two aspects. One is the international voluntary committee system by which the need for change is communicated to the International Federation for Documentation (FID) and its Central Classification Committee. The other is the technical difficulty of maintaining and disseminating up-to-date classification schedules in many languages. The latter reflects back on the former in that it is often difficult even for a revision committee to obtain the most recent and complete version of the schedules for its

specialty. In project report number 3, we summarized the problem in quantitative terms, as follows:

> "Quantitatively speaking, if we assume that a reasonable goal for the UDC is to be available in a complete form in a total of twenty languages, the UDC would be a file of approximately the following size:
>
> | | |
> |---|---|
> | 125,000 | records in a full edition |
> | 20 | languages |
> | 2,500,000 | records |
> | 200 | characters per record, including UDC number, heading, cross references, and alphabetic index entries |
> | 500,000,000 | characters in total UDC file |
>
> The problem of the UDC would be to keep such a file up to date, disseminate changes to users rapidly, and select various portions of the file to be printed periodically, according to managerial decisions as to needs for new full editions, abridgements, and special subject editions. A further complication of no small magnitude is that the encoding and display mechanisms must provide for all of the orthographic forms commonly used by the twenty (or more) languages, as well as mathematical symbols."

The technical aspect of the problem can be solved, as demonstrated by the work reported in project reports 2 and 3 and by the work of the staff of the Zentralstelle für maschinelle Dokumentation in Germany. An effort is needed on the part of FID or its national members to develop a permanent base and an expert staff to proceed toward the solution.

4.4. Seminar on UDC in a Mechanized Retrieval System. One of the most concrete, yet difficult to document, results of this project is an operational information retrieval system capable of using UDC as the indexing language (the Combined File Search System). Anticipating requests from potential users for instructions on how to operate the system, we suggested a one-week seminar at which the entire process could be reviewed in step-by-step detail, with demonstrations on a computer. The suggestion was accepted by

the FID Classification Research (FID/CR) and Central Classification (FID/CCC) Committees. The Danish Centre for Documentation, located at the Technical Library of Denmark, has agreed to act as a secretariat and to provide meeting facilities. The North European University Computer Center likewise agreed to provide computer time. The seminar is scheduled tentatively for September 2-6, 1968.

5. Acknowledgements. The AIP/UDC Project benefitted from the splendid cooperation of many individuals and organizations, spread over the United States, Canada, and many parts of Europe. While their names are too numerous to list individually, their services as consultants, advisors, subcontractors, and friends of the project are nonetheless sincerely appreciated.

APPENDIX 1

Bibliography of AIP/UDC Project Reports, with Abstracts

1. Freeman, Robert R., Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System: An Introduction, New York, American Institute of Physics, Report AIP/DRP UDC-1, October 1, 1965. NSF Grant GN-433.

The report describes the five areas of activity which lead toward the aim expressed in the title: (1) to develop a complete English-language version of UDC in both hierarchical and alphabetical arrangement in machine-readable form; (2) to develop techniques for automatic file maintenance and photo-composition of UDC editions; (3) to develop a computer-based reference retrieval system which uses UDC as its indexing language; (4) to collect a set of UDC-indexed document files in machine-readable form in various subject areas; and (5) to conduct tests with the aid of experimental user groups which will lead to an evaluation of the UDC in the desired context. Data are also given on the organization of the project.

2. Freeman, Robert R., Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System: Progress Report for the Period July 1, 1965 - January 21, 1966, New York, American Institute of Physics, Report AIP/DRP UDC-2, February 1, 1966. NSF Grant GN-433.

The report reviews activities involving collection of English, French, and German schedules of the Universal Decimal Classification (UDC), translation of some schedules, further development of a mechanized (IBM 1401) UDC file maintenance system, experiments with automatic alphabetic indexing of UDC schedules, automatic typesetting and composition of UDC schedules, selection of equipment and rules for keyboarding the UDC into machine readable form, and initial steps toward collections of UDC-indexed documents and a retrieval system for test and evaluation purposes. Detailed appendices deal with considerations of creating machine-readable UDC records on punched-paper tape for subsequent computer processing.

3. Freeman, Robert R., Modern Approaches to the Management of a Classification, Report AIP/UDC-3 under National Science Foundation Grant GN-433, New York, American Institute of Physics, October 1, 1966. Presented at the Seminar on UDC and Mechanization at the 32nd Conference of the International Federation for Documentation, the Hague, September 20, 1966. Also published as "The Management of a Classification: Modern Approaches Exemplified by the UDC Project of the American Institute of Physics," Journal of Documentation, 23(4), 304-320 (December, 1967).

The report views the problem of managing a classification, such as the Universal Decimal Classification (UDC), as an example of the broader class of problems known in the system analysis and data processing field as "file management". The characteristics of file management are listed and related specifically to the UDC. The uses of data processing equipment for the creation, maintenance, manipulation and display of files are discussed. The development of a prototype file management system for the UDC is reviewed. Appendices illustrate the progress of the project and summarize the present status of the UDC in the English language.

4. Russell, Martin, and Freeman, Robert R., Computer-Aided Indexing of a Scientific Abstracts Journal by the UDC with UNIDEK: a Case Study, Report AIP/UDC-4 under National Science Foundation Grant GN-433, New York, American Institute of Physics, April 1, 1967.

This paper is a case study of the adoption by Geoscience Abstracts of UNIDEK, a novel computer-compiled systematic subject index based on the Universal Decimal Classification (UDC) of the International Federation for Documentation (FID). Events leading to a decision to adopt the system, some theory of indexes, problems involved in conversion, and some of the results achieved are reviewed.

5. Freeman, Robert R. and Pauline Atherton, File Organization and Search Strategy Using the Universal Decimal Classification in Mechanized Reference Retrieval Systems, Report AIP/UDC-5 under National Science Foundation Grant GN-433, New York. American Institute of Physics, September 15, 1967. Presented at the FID/IFIP Conference on Mechanized Information Storage, Retrieval, and Dissemination, Rome, June 15, 1967. Published in Proceedings of the Conference, North Holland Publishing Co., (forthcoming).

Starting from a model of contemporary mechanized retrieval systems and the characteristics of indexing languages used therein, the authors develop a rational basis for use of the Universal Decimal Classification (UDC) in this context. Practical design considerations for the use of UDC in a mechanized retrieval system are discussed. Examples are reported of the use of UDC as the indexing language with the Combined File Search System, an existing retrieval system for the IBM 1401, used by several large information centers in the United States. Finally, the authors discuss how UDC might be used as a query language in a typical retrieval system of the near future in which the user interacts directly with the computer-stored document reference file.

The authors conclude that it is technically feasible to use UDC in mechanized retrieval systems and that, under certain conditions, it may be desirable. Some of these conditions are the existence of large files already indexed by UDC, staff already trained for its use, and extensive international use or exchange of materials of the system.

6. Freeman, Robert R., Evaluation of the Retrieval of Metallurgical Document References Using the Universal Decimal Classification in a Computer-Based System, Report AIP/UDC-6 under National Science Foundation Grant GN-433, New York, American Institute of Physics, April 1, 1968.

A set of twenty-five questions were processed against a computer-stored file of 9,159 document references in the field of ferrous metallurgy, representing the 1965 coverage of the Iron and Steel Institute (London) information service. A basis for evaluation of system performance characteristics and analysis of system failures was provided by using questions which had previously been processed by the American Society for Metals against a data base which contained many of the same documents. The Cuadra-Katter model for describing the system evaluation environment was used. The results, which were highly satisfactory, led to observations and recommendations which

contrast the requirements for class definition, indexing policy, and search strategy between manual and computer-based systems which use UDC.

7. Freeman, Robert R. and Pauline Atherton, AUDACIOUS - an Experiment with an On-Line, Interactive Reference Retrieval System Using the Universal Decimal Classification as the Index Language in the Field of Nuclear Science, Report AIP/UDC-7 under National Science Foundation Grant GN-433. New York, American Institute of Physics, April 25, 1968.

The report describes an experimental system for remote direct access to files of computer-stored information which has been indexed by the Universal Decimal Classification (UDC). The data base for the experiment consisted of references from a single issue of Nuclear Science Abstracts. The Special Subject Edition of UDC for Nuclear Science and Technology was also stored in the computer so that users could discover how to translate their questions from natural language to logical statement containing UDC numbers.

The authors conclude that the technical feasibility of use of existing classification and indexing tools, such as UDC, has been demonstrated. However, detailed attention to all facets of man-machine communication is a necessity if systems are to be designed which will be voluntarily used. AUDACIOUS is reviewed and criticized from this point of view.

Finally, the authors conclude that the use of UDC in an on-line, interactive system may have important ramifications for the development of international information networks. Conversion tables (schedules) already exist which would allow speakers of many languages to search files indexed by UDC without regard to national or linguistic boundaries.

8. Atherton, Pauline, Donald W. King, and Robert R. Freeman, Evaluation of the Retrieval of Nuclear Science Document References Using the Universal Decimal Classification in a Computer-Based System, Report AIP/UDC-8 under National Science Foundation Grant GN-433, New York, American Institute of Physics, May 1, 1968.

A single issue of Nuclear Science Abstracts, containing about 2,300 abstracts, was indexed by UDC, using the Special Subject Edition of UDC for Nuclear Science and Technology. The descriptive cataloging and UDC-indexing records formed a computer-stored data base. A systematic random sample of 500 additional abstracts, taken from a collection of about 196,000, was also indexed by UDC. An experimental design was developed such that the potential results of retrieval tests with the full collection could be inferred from actual results obtained from the two smaller data bases.

Sixty questions were collected from nuclear science research organizations in North America and Europe. Two search analysts, neither of whom was familiar with the policies and practices of the indexers, formulated logical search statements with UDC numbers. The resulting queries were processed against the UDC-indexed data bases. They were also processed by two other information services. Twelve questions, a subset of the original sixty, were chosen for more detailed analysis. The results are presented in the report.

9. Freeman, Robert R., and Pauline Atherton, Final Report of the Research Project for the Evaluation of the UDC as the Indexing Language for a Mechanized Reference Retrieval System, Report AIP/UDC-9 under NSF Grant GN-433. New York, American Institute of Physics, May 1, 1968.

The background, objectives, and accomplishments of the project are reviewed briefly. Specific areas discussed are English language UDC schedules, a computer-based UDC file management system, data bases for retrieval experiments, batch-process and on-line, interactive information retrieval systems, and retrieval system evaluation. The conclusions deal with the usefulness of the UDC for mechanized retrieval systems, needed research, needed organizational effort, and a proposed international seminar. Several appendices summarize the current state of the UDC in English and the availability of magnetic tape and microfilm files developed by the project.

APPENDIX 2

AIP/UDC Project Magnetic Tapes and Other Tapes Containing AIP/UDC Project Materials

A. Inquiries concerning the following tapes should be directed to the Information Division, American Institute of Physics, 335 East 45th Street, New York, New York 10017.

Tape	Tracks	Density	Contents
AIP-1	9	800	Geology document file, reel 1 of 2
AIP-2	9	800	Geology document file, reel 2 of 2
AIP-3	9	800	UDC English Language Master File (see Appendix 5)
AIP-4	7	800	UDC English Language Master File
AIP-5	9	800	Geology descriptor file
AIP-6	9	800	Metallurgy descriptor file
AIP-7	7	556	UDC Special Subject Edition for Metallurgy
AIP-8	7	556	Combined File Search System Program Tape
AIP-9	9	800	Metallurgy document file
AIP-10	7	556	UDC numbers from German Medium Edition (numbers only, no text)
AIP-11	-	-	-
AIP-12	9	800	UDC Reverse Cross reference file (see Appendix 5)
AIP-13	7	800	UDC Abridged Building Classification - 1968
AIP-14	9	800	Nuclear Science document file
AIP-15	9	800	Nuclear Science descriptor file
AIP-16	7	800	Nuclear Science document file
AIP-17	7	800	Nuclear Science descriptor file
AIP-18	9	800	Combined File Search System - System Tape

B. Inquiries concerning the following tapes should be directed to the American Meteorological Society, P.O. Box 1736, Washington, D.C. 20013.

Tape	Tracks	Density	Contents
AMS-1	9	800	Meteorology document file
AMS-2	9	800	Meteorology descriptor file
AMS-3	9	800	Oceanography document file
AMS-4	9	800	Oceanography descriptor file
AMS-5	9	800	Antarctic document file
AMS-6	9	800	Antarctic descriptor file

APPENDIX 3

SUMMARY OF THE STATUS OF THE UNIVERSAL DECIMAL CLASSIFICATION IN ENGLISH, INCLUDING PROGRESS BY THE AIP/UDC PROJECT THROUGH 31 December 1967

KEY TO CODES USED IN TABLE

1 = Published by British Standard Institution [BS1] (except for Special Subject Editions)

2 = Unpublished manuscript

2a = Manuscript reported available at BS1, but not in possession of AIP/UDC Project Staff

3 = Section entirely covered in published Special Subject Edition

4 = English Medium Edition text derived by condensation and editing of published full edition of manuscript through comparison with German Medium Edition* and Extensions and Corrections (E+C)

4a = Unofficial translation completed by AIP/UDC Project Staff

5 = Manuscript not yet completed

All check-marked entries in the following table correspond to records included in the AIP/UDC Project English Language Master File (Appendix 5)

All entries bearing the code number 4 in the following table correspond to English Medium Edition manuscript included in Appendix 4.

---

*Deutscher Normenausschuss, Dezimalklassifikation: DK-Handausgabe, Band 1 Systematische Tafeln, Beuth-Vertrieb GmbH, Berlin. and Köln, 1967.

SUMMARY OF THE STATUS OF THE UNIVERSAL DECIMAL CLASSIFICATION IN ENGLISH INCLUDING PROGRESS BY THE AIP/UDC PROJECT THROUGH 31 DECEMBER, 1967

UDC SECTION		FULL EDITION		MEDIUM EDITION			ABRIDGED EDITION	
		Current Publication Status	Converted to Machine-Readable Form	Current Manuscript Status	Converted to Machine-Readable Form	Columns in Equivalent German Edition	Current Publication Status	Converted to Machine-Readable Form
0	Generalities	1	✓	4	✓	14	1	✓
1	Philosophy, Psychology	2a		4	✓	12	1	✓
2	Religion	2a		4	✓	15	1	✓
3/30	Sociology	2	✓	4	✓	3	1	✓
31	Statistics	2	✓	4	✓	2	1	✓
32	Political Science	2	✓	4	✓	6	1	✓
33	Economics	2	✓	4	✓	17	1	✓
34	Law			4	✓	27	1	✓
35	Public Administration			4	✓	16	1	✓
36	Social Welfare			4	✓	8	1	✓
37	Education	3(ED)	✓	4	✓	9	1	✓
38	Commerce			4	✓	3	1	✓
39	Social Anthropology			4	✓	3	1	✓
5/51	Mathematics	1	✓	4		9	1	✓
52	Astronomy	1	✓	4		11	1	✓
53	Physics	1	✓	4		30	1	✓
54	Chemistry	1	✓	4		37	1	✓
55	Geology, Meteorology	1	✓	4		16	1	✓
56	Paleontology	1	✓	4		1	1	✓
57	General Biology	1	✓	4		9	1	✓
58	Botany	1	✓	4		11	1	✓
59	Zoology	1	✓	4		16	1	✓

UDC SECTION		FULL EDITION		MEDIUM EDITION			ABRIDGED EDITION	
		Current Publication Status	Converted to Machine-Readable Form	Current Manuscript Status	Converted to Machine-Readable Form	Columns in Equivalent German Edition	Current Publication Status	Converted to Machine-Readable Form
6/60	Appl. Sci. & Tech.			4	✓	½	I	✓
61	Medicine	2a		4	✓	49	I	✓
62/62-9	Engineering & Technology-gen.	2a		5		11	I	✓
620.1	Materials Testing	3 (ME)	✓	5		4	I	✓
621.039	Nuclear Engineering	3 (NS)	✓	5		4	I	✓
621.1/.2	Steam & Water Power	2a		5		3	I	✓
621.3	Electrical Engineering	I	✓	5		20	I	✓
621.4	Combustion Engines	2a		5		3	I	✓
621.5	Pneumatic Energy	2a		5		2	I	✓
621.6	Fluid Distribution	2	✓	5		3	I	✓
621.7	Workshop Practice	3 (ME)	✓	5		12	I	✓
621.8	Power Transmission	2a		5		8	I	✓
621.9	Tools. Machining	3 (ME)	✓	5		8	I	✓
622	Mining Engineering	I	✓	5		11	I	✓
623	Military Engineering	I	✓	5		9	I	✓
624	Civil Engineering	2	✓	5		11	I	✓
625	Railways & Highways	2	✓	5		9	I	✓
626	Hydraulic Engineering	2	✓	5		4	I	✓
627	River, Port, Harbor, Coastal Eng.	2	✓	5		6	I	✓
628	Public Health Engineering	2	✓	5		9	I	✓
629	Transport Engineering	2(E+C)	✓	5		10	I	✓
63/635	Agriculture	2	✓	5		23	I	✓
636	Stockbreeding	2	✓	5		4	I	✓
637/639	Animal Husbandry	2	✓	5		8	I	✓
64	Home Economics	2a		5		6	I	✓
651	Office Management	2	✓	5		3	I	✓
654	Telecommunication	2	✓	5		2	I	✓
655	Printing & Publishing	2	✓	5		2	I	✓
656	Transport & Postal Svc.	2	✓	5		8	I	✓
657	Accounting	2	✓	5		3	I	✓
658	Management Techniques	2	✓	5		8	I	✓
659	Publicity, Information	2	✓	5		3	I	✓

UDC SECTION		FULL EDITION		MEDIUM EDITION			ABRIDGED EDITION	
		Current Publication Status	Converted to Machine-Readable Form	Current Manuscript Status	Converted to Machine-Readable Form	Columns in Equivalent German Edition	Current Publication Status	Converted to Machine-Readable Form
66/661	Chem. Eng.	2	✓	5		19	1	✓
662	Explosives, Fuels	2	✓	5		10	1	✓
663	Beverages, Stimulants	2	✓	5		6	1	✓
664	Food and Preservation	2	✓	5		8	1	✓
665	Oils, Fats, Waxes	2a		5		5	1	✓ *
666	Glass & Ceramics	2a		5		12	1	✓
667	Paints & Dyes	2a		5		5	1	✓
668	Misc. Organics	2	✓	5		3	1	✓
669	Metallurgy	3 (ME)	✓	5		15	1	✓
67/673	Processable Materials	2a		5		3	1	✓
674	Timber-Woodwork			5		2	1	✓
675	Leather			5		2	1	✓
676	Paper	2a		5		5	1	✓
677	Textiles	2	✓	5		13	1	✓
678	Macromolecular Materials	1	✓	5		5	1	✓
679	Stone Industry	1	✓	5		1	1	✓
68/681.2	Precision Instruments	2a		5		3	1	✓
681.3	Data Processing Equipment	2 (E+C)	✓	5		5	1	✓
681.4	Optical Instruments	2a		5		1	1	✓
681.6	Reproducing Machines	2	✓	5		5	1	✓
681.8	Technical Acoustics	2	✓	5		2	1	✓
681.9	Engraving Machines			5		½	1	✓
682/683	Iron-working	2	✓	4a	✓	3	1	✓
684	Furniture	2 (E+C)	✓	4a	✓	3	1	✓
685	Outdoor, Travel, and Sports Equipment			4a	✓	3	1	✓
686	Bookbinding, Stationery			4a	✓	2	1	✓
687	Clothing	2a		4a	✓	2	1	✓
688/689	Decorative Goods, Hobbies, Handcrafts	2a		4a	✓	2	1	✓
69	Building Industry	1	✓	3 (ABC)		11	1	✓

UDC SECTION		FULL EDITION		MEDIUM EDITION			ABRIDGED EDITION	
		Current Publication Status	Converted to Machine-Readable Form	Current Manuscript Status	Converted to Machine-Readable Form	Columns in Equivalent German Edition	Current Publication Status	Converted to Machine-Readable Form
7.0	Art	2	✓	4	✓	3	1	✓
71	Physical Planning	2	✓	4	✓	4	1	✓
72	Architecture	2	✓	4	✓	4	1	✓
73/76	Sculpture, Drawing, Painting, Engraving			4	✓	4	1	✓
77	Photography	2	✓	4	✓	13	1	✓
78	Music	2a		4	✓	3	1	✓
79/795	Entertainment, Theatre, Games			4	✓	2	1	✓
796/799	Sports	3 (SP)	✓	4	✓	3	1	✓
8/81	Language and Linguistics	2 (E+C)	✓	4	✓	6	1	✓
82/89	Literature	2 (E+C)	✓	4	✓	3	1	✓
9	Geog., History, Biography			4	✓	6	1	✓
	AUXILIARIES							
=	Language	1	✓	5		2	1	✓
(0..)	Form of Work	1	✓	5		5	1	✓
(1/9)	Place	1	✓	5		49	1	✓
(=...)	Race, Nationality	1	✓	5		1/2	1	✓
"..."	Time	1	✓	5		2	1	✓
.00	Point of View	1	✓	5		3	1	✓
	SPECIAL SUBJECT EDITIONS							
Nuclear Science (NS)		1	✓					
Metallurgy (ME)		1	✓					
Education (ED)		1	✓					
Building (ABC)		1						
Standardization		1						
Polar Regions		1						
Sport and Physical Education (SP)		1	✓					

APPENDIX 4

UDC ENGLISH MEDIUM EDITION MANUSCRIPT. CLASSES 0, 1, 2, 3, 5, 7, 8 and 9

Introduction. This appendix includes the full manuscript for the main classes of an English language medium edition of UDC with the exception of class 6, the largest class. The manuscript was prepared under the supervision of Mr. Geoffrey Lloyd at the headquarters of the International Federation for Documentation (FID) in the Hague, with the support of funds transferred to FID by the AIP/UDC Project with permission of the National Science Foundation.

The raw material for the manuscript was (1) an earlier version of the AIP/UDC Project English Language Master File which is included in this report as Appendix 5, (2) recent Extensions and Corrections to the UDC, and (3) the German and French Medium Editions of UDC. Of the total manuscript reproduced here, all but class 5 was subsequently entered into the English Language Master File.

By arrangement with FID, the British Standards Institution has tentatively agreed to publish the English Medium Edition. Actual publication will be contingent upon completion of UDC class 6, the auxiliary classes, and an alphabetic index to the edition. In the interests of making the manuscript available for the use of researchers and teachers of library science, the incomplete manuscript, amounting to 687 typewriten pages, has been reporduced as described below.

Availability. Owing to the large number of pages in the manuscript, only the introductory pages of this appendix are reproduced here. The full manuscript will be made available in microform according to the arrangements detailed in appendix 5.

APPENDIX 5

THE UDC ENGLISH LANGUAGE MASTER FILES

PART 1: THE CLASSIFICATION SCHEDULE

PART 2: REVERSE CROSS REFERENCES

Introduction. This appendix includes the complete,merged set of UDC schedules accumulated by the AID/UDC Project from 1965-1967. The file, which contains 110,759 records, is stored on magnetic tape recorded at a density of 800 cpi in IBM BCD code. Both 7 and 9-track versions exist. The reverse cross reference file, which contains 24,140 records is also stored on magnetic tape.

Each record in the schedule includes the UDC number, a code which represents the source of the record, and the English language equivalent of the UDC number. Many records also contain cross references (Cf...) and scope notes which serve to further define or delimit a concept.

Cross references were identified and recorded in a separate file, as well as in the UDC schedule. The former was sorted and printed in the order of the UDC numbers referred to by a given record. The resulting listing appears here as Part 2, the reverse cross reference file. The source codes which appear in both parts of this appendix are explained below.

Source Edition Codes Used in the Classification Schedule and the Reverse Cross Reference File

Code	Meaning
ENAB61	Abridged Edition, 1961, complete
ENED65	Education Edition, 1965
ENFU--	Full Edition, unpublished ms.
ENFU43	Full Edition, 1943 (partial)
ENFU54	Full Edition, 1954 (partial)
ENFU55	Full Edition, 1955 (partial)
ENFU58	Full Edition, 1958 (partial)
ENFU64	Full Edition, 1964 (partial)
ENME67	Medium Edition, 1967 (not yet published)
ENMT64	Metallurgy Edition, 1964
ENNS64	Nuclear Science Edition, 1964
ENSP64	Sport and Physical Education Edition, 1964

(EN indicates an English language schedule. Two letters serve as an edition code and two digits identify the year of publication).

Availability. Part one of this appendix contains 2,659 pages, while part two contains 409 pages. Owing to this large size, only the introductory section of the appendix is reproduced in the final report. The following arrangements have been made in order to assure availability of the full appendix to persons interested in its use for research and teaching purposes.

1. The appendix has been recorded on 16 mm negative reel microfilm. Arrangements for obtaining copies may be made by contacting the Information Division, American Institute of Physics, 335 East 45th Street, New York, New York 10017.

2. The final report including this appendix will be available on microfiche through the Educational Resources Information Center (ERIC) of the U.S. Office of Education. The final report will be processed by the ERIC Clearinghouse on Library and Information Sciences, 2122 Riverside Avenue, Minneapolis, Minnesota 55404. After announcement in the ERIC abstracting service, Research in Education, the appendix will be available through the ERIC Document Reproduction Service, the National Cash Register Co., 4936 Fairmont Avenue, Bethesda, Maryland 20014.

3. Since the appendix will be available through the ERIC system, it will not be made available through the Clearinghouse for Federal Scientific and Technical Information.

4. Arrangements for obtaining copies of the tapes may be made by contacting the Information Division, American Institute of Physics, 335 East 45th Street, New York, New York 10017.